Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Research Note—Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities

Gerald C. Kane, Sam Ransbotham

Information Systems Research
Vol. 27, No. 2, June 2016, pp. 424–439
ISSN 1047-7047 (print) | ISSN 1526-5536 (online)

http://dx.doi.org/10.1287/isre.2016.0622

Research Note

Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities

Gerald C. Kane, Sam Ransbotham
Boston College, Chestnut Hill, Massachusetts 02467
{gerald.kane@bc.edu, sam.ransbotham@bc.edu}

The 15-year history of collaboration on Wikipedia offers insight into how peer production communities create knowledge. In this research, we combine disparate content and collaboration approaches through a social network analysis approach known as an affiliation network. It captures both how knowledge is transferred in a peer production network and also the underlying skills possessed by its contributors in a single methodological approach. We test this approach on the Wikipedia articles dedicated to medical information developed in a subcommunity known as a WikiProject. Overall, we find that the position of an article in the affiliation network is associated with the quality of the article. We further investigate information quality through additional qualitative and quantitative approaches including expert coders using medical students, crowdsourcing using Amazon Mechanical Turk, and visualization using network graphs. A review by fourth-year medical students indicates that the Wikipedia quality rating is a reliable measure of information quality. Amazon Mechanical Turk ratings, however, are a less reliable measure of information quality, reflecting observable content characteristics such as article length and the number of references.

Keywords: social media; information quality; network analysis
History: Natalia Levina, Senior Editor; Jonathon Cummings, Associate Editor. This paper was received on May 20, 2014, and was with the authors 12 months for 2 revisions. Published online in *Articles in Advance* April 4, 2016.

Information Quality in Peer Production Communities

Wikipedia can be a source of insight for studying collaboration in peer production communities (Kane and Fichman 2009). Since Wikipedia has preserved nearly 15 years of collaborative activity, it can serve as a resource for understanding other peer production communities with a more limited history. A complete review of all of the research literature investigating Wikipedia is daunting. Indeed, a recent review identifies nearly 500 research studies involving Wikipedia (Okoli et al. 2012). In this paper, we focus primarily on the factors associated with information quality on Wikipedia.

Previous research suggests two main approaches to studying information quality on Wikipedia—content and collaboration (e.g., Keegan et al. 2013, Warncke-Wang et al. 2013). *Content* approaches focus on whether the composition of the article—such as article length, number of references, number of headings, number of links, and number of images—is related to the overall quality of the article (e.g., Stvilia et al. 2008). *Collaboration* studies focus on whether the nature of interaction between contributors is somehow related to the quality of the information the community produces (e.g., Kittur and Kraut 2008, Kittur et al. 2009). Each approach has its own strengths and limitations; greater unification between the content and collaboration approaches for understanding information quality may be beneficial (Keegan et al. 2013). Researchers have used social network analysis (SNA) in both approaches to information quality and it offers a promising route to unify them. We argue that a conceptualization known as an *affiliation network*—which treats articles as nodes and contributors as the ties that connect them—overcomes many of the drawbacks of previous studies that study information quality on Wikipedia (Ransbotham et al. 2012).

To test how well an affiliation network predicts information quality on Wikipedia, we examine the article quality of 16,244 articles built through 2,677,397 revisions by 147,362 distinct contributors to Wikipedia's Medicine WikiProject, from February 2001 to August 2011. We first evaluate the information quality by using a team of medical experts to validate the reliability of Wikipedia's crowdsourced quality measures as a proxy for the quality of the underlying medical information contained in the articles. We then find support for our hypotheses that the quality of information found in a peer production community is related to the affiliation network structure. We also explore crowdsourced Amazon Mechanical Turk ratings of information quality, finding that these ratings primarily identify differences in content characteristics between complete articles and article fragments, rather than serving as a robust measure of information quality.

Content Approaches to Information Quality

The content approach to assessing information quality on Wikipedia primarily examines how well characteristics of the article predict the quality of the information found in it (e.g., Stvilia et al. 2008). For example, simple word count (Blumenstock 2008), writing style (Lipka and Stein 2010), particular types of language choices (Xu and Luo 2011), and the type of content added to or deleted from an article (Liu and Ram 2011) are all associated with information quality. Perhaps in the most comprehensive content approach to information quality, Warncke-Wang and colleagues (2013) developed an approach of information quality that utilized only content from the article to predict information quality. They argue that such a content-based approach is superior because it provides actionable information to the contributors that they can use to improve the article. Although the "actionable" nature of the content approaches may offer benefits, it may also be viewed as a liability, because it can be difficult to assess if observed correlations are causal. Once contributors know the quality metrics, production communities can target those metrics as the primary goal itself, and they may cease to function as a latent measure of quality. For example, if contributors know that the best articles are roughly 5,000 words in length, they may add filler or delete valuable content just to hit the word count targets, not to improve the content.

The content approach to understanding information quality has used SNA; understanding articles are a part of a wider *content network*. Under this conceptualization, articles are the nodes of the network, connected by hyperlinks. This conceptualization is similar to the initial form of Google's PageRank algorithm that assessed the importance of webpages on the Internet (Brin and Page 1998). In analyzing the content network in this way, such networks exhibit a power-law feature in which a few select articles were heavily linked to but most articles were not (Capocci et al. 2008). Articles that enjoyed these high numbers of links were more likely to be contributed to by contributors who were not intimately familiar with the content (Halatchliyski et al. 2010). Yet, little research connects the features of this *content network* with information quality, perhaps because the hyperlinks on Wikipedia do not capture the same type of information as on the Internet in general. Since many pages are associated with the same topic on the Internet, the prevalence of incoming hyperlinks helps select the best content among these many choices. Wikipedia, however, only hosts one article per topic, potentially rendering incoming hyperlinks a measure of topical interest instead of information quality.

Collaboration Approaches to Information Quality

The collaboration approach to understanding information quality on Wikipedia assesses how the patterns of interaction between contributors affect the quality of the information they create (e.g., Kittur and Kraut 2008). The collaboration approach often uses SNA to examine how well the contributors to a particular article work together, which we call the *edit network.* This approach examines the network of how contributors (nodes) contribute to a particular article (ties). A considerable amount of edit network research examines the role of conflict in the collaboration process, which results in lower quality articles (Brandes et al. 2009, Maniu et al. 2011, Jurgens and Lu 2012). Others conceptualize the collaboration network differently, examining the network between contributors (nodes) through their interactions with one another on talk pages (ties), which we call the *talk page network* (Massa 2011). The structure of these talk page networks captures whether these contributors communicate with each other while working on an article. This network is associated with how efficiently contributors improve the information in an article and moves it from one quality designation to the next (Nemoto et al. 2011).

A pure collaboration approach also has some drawbacks. Few collaboration studies control for content characteristics of the article. This omission is particularly problematic given the considerable research demonstrating the association between content characteristics and information quality. As a result, collaboration studies have trouble distinguishing whether particular networks are generating better content or simply more of it. The content also influences the type of collaboration that occurs around it. For example, different article topics and genres exhibit distinct collaboration network structures (Iba et al. 2010, Keegan et al. 2013). Articles also often exhibit different collaboration patterns at different stages in its production lifecycle, patterns that are associated with article quality milestones (Kane et al. 2014b). The article serves as a coordinating mechanism for determining what type of collaboration is needed at a particular stage. Considering this coordinating role, it may also be important to account for the content itself as an integral part of the collaboration network.

Another drawback of the collaboration research stream is that most studies examine only the collaboration occurring within the context of a given article. However, contributors typically do not work only on a single article; instead, they often work on multiple articles (Wang et al. 2013). A shared contributor serves as a conduit for transferring information and experience—consciously or not—between the two articles. In this case, the work on one article could influence the development of another. Additionally, a contributor's work on other articles may provide insight about the knowledge she possesses. These cross-article connections have been largely overlooked in the studies connecting collaborative patterns to information quality, yet they may prove important.

Affiliation Networks As a Unified Approach to Information Quality

We argue that an integrated approach to information quality on Wikipedia should account for (1) the content characteristics of the article, (2) the articles as a part of the collaboration network, and (3) collaboration activity that occurs across articles. A particular type of network, known as an *affiliation network*, can accomplish these objectives. An affiliation network incorporates both contributors and articles into the same network, treating the articles as nodes and the contributors as ties that connect them. The use of affiliation networks to study the characteristics of groups dates back to a seminal study of the participation of southern women in shared social events, conceptualizing the network of events connected by the women who participated in each (Davis et al. 1941). Since then, affiliation networks show how people relate to one another through their membership in shared events (members and events), corporate board interlocks (members and boards), political affiliations among legislators (senators and bills), opinions of Supreme Court justices (judges and opinions), scientific communities (contributors and papers), and many other types of clustered social organizations (Carrington et al. 2005).

Membership in a shared group is often regarded as a distinct type of tie in a social network, through which members are exposed to similar information and experiences, even if they do not interact directly (Borgatti et al. 2009, Kane et al. 2014a). Similar rationale has been used in exploring the role of affiliation networks in other types of peer production communities (e.g., open source software (OSS)). For example, the success of an OSS project is related to the affiliation network connections among project leaders (Grewal et al. 2006) and the project's position in the project-developer network (Fershtman and Gandal 2011, Singh and Tan 2011, Singh et al. 2011a).

Peer production communities on Wikipedia, however, differ from OSS in important ways. Because of its widespread appeal, easy access, and low barriers to entry, Wikipedia communities exhibit higher membership fluidity than OSS projects—defined here as members leaving one community and joining another (Faraj et al. 2011). This higher membership fluidity results in considerably different affiliation network structures than OSS, which may have important and unknown effects on its relationship to the quality of information created by a community. Higher fluidity yields larger average community size as people join and leave with greater frequency. It may also yield a denser affiliation network, as people often join another peer production community when they leave one. This denser affiliation network structure could alter the effect the affiliation network has on information quality. We expect that the affiliation network will have a *positive* effect on the information quality created by the community. Collaborators that move more readily from one group to another may have more opportunities to transfer knowledge gained in one group to another (O'Dell and Grayson 1998). They are also more exposed to the work of other contributors, and contributors often emulate beneficial behaviors in their work on other articles (Zhu et al. 2012).

Other research from both Wikipedia and social network literatures indicates that the affiliation network structure on information quality is an open question. The different network structures resulting from higher fluidity could also result in a *negative* effect of the network on information quality. Negative attitudes and experiences spread through the network just as easily as positive ones and often have a more powerful effect on outcomes (Labianca and Brass 2006). Conflict is rampant on Wikipedia (Arazy et al. 2011), and negative experiences with previous collaborators spill over to collaborations on other articles (Rad and Barbosa 2012). Thus, a denser affiliation network may expose communities to more negative information and behaviors, adversely affecting quality. These different network structures that result from higher fluidity could also simply serve to revert information quality to a *mean state*. Too much fluidity in communities that have created high-quality articles is also associated with the deterioration of the quality of the information the community created (Ransbotham and Kane 2011). As contributors join and leave multiple communities, they may also learn less from their participation in each community, have less knowledge to transfer, and not develop bonds with other contributors that facilitate production. When contributors bring different but not better content, more contributions reverts information quality to average levels (March 1991).

How Affiliation Networks Affect Information Quality

Given that we expect a net positive effect of the affiliation network on collaboration, we hypothesize that three aspects of the affiliation network affect the quality of information on Wikipedia: (1) the number of contributors, (2) the quality-weighted degree centrality of the article in the affiliation network, and (3) the eigenvector centrality of the article in the affiliation network.

The total number of contributors to a Wikipedia article can be the single biggest predictor of information quality (Wilkinson and Huberman 2007). A greater number of contributors have a greater total amount of information, effort, and energy available to create and improve the article (Friedkin 1982). More contributors also make it likely that someone will identify errors in the article. A similar principle applies in OSS communities through a popular mantra—with enough eyeballs, all bugs are shallow (Raymond 1999). People also may only contribute if they feel they can improve the article (Preece and Schneiderman 2009), so the number of contributors may also be associated with the total number of contributions to it and reflect how well developed the content is.

However, too many contributors create problems that limit the value of new members past a certain point. Too many contributors give rise to conditions of information overload (Hiltz and Turoff 1985), which reduces a group's ability to organize information effectively and makes it more difficult to collaborate (Kuk 2006, Kittur and Kraut 2010). New contributors add overhead costs because the community must assess the potential value of the information possessed by the contributor. Increasing information availability also makes it difficult to pay attention to and identify which information is most valuable (Hansen and Haas 2001). It also makes it more difficult to get attention for one's contribution, a motivation for many people to contribute to online communities, which could discourage new contributors from participating (Butler 2001). The preceding arguments suggest a curvilinear relationship between the number of contributors and the quality of an information artifact. That is, additional contributors are beneficial, up to a certain point. Quality improves if enough contributors are attracted to generate sufficient collaborative activity, but begins to decrease at the point when it overloads contributors. We hypothesize the following:

Hypothesis 1. *The number of contributors to an information artifact has a curvilinear relationship with information quality, such that a moderate number of contributors create the highest quality information.*

Additional contributors are only valuable to the extent that they possess new knowledge not already available to the community (Asvanund et al. 2004). The *diversity* of knowledge reflects the amount of novel and nonoverlapping knowledge possessed by a production community. Diversity of knowledge possessed by contributors is a precursor of collective intelligence (Surowiecki 2004). For example, a greater diversity of knowledge held by respondents to a listserv query is more likely to result in a useful response (Constant et al. 1996). The structure of social networks can reflect knowledge diversity. Weak-tie networks are more likely to result in valuable outcomes in certain situations, because they provide access to different information sources (Granovetter 1973). Extending this logic, the structure of the affiliation network may be an indicator of knowledge diversity. We hypothesize that the number of other articles that editors contribute to—an affiliation network structure known as *degree centrality*—reflects the knowledge diversity available to a community and positively relates to quality.

Degree centrality in the affiliation network reveals information quality in two ways (Podolny and Baron 1997). Ties serve as direct conduits between nodes. Contributors can directly transfer content (O'Dell and Grayson 1998) or behavioral knowledge (Zhu et al. 2012) from one article to another, increasing the resources available to the peer production community. When contributors work on more and different articles, they can transfer knowledge or experience from more sources. Networks also serve as signals of the latent characteristics possessed by its members' characteristics. Here, the affiliation network reflects the underlying knowledge possessed by contributors. People with one type of knowledge (e.g., medical knowledge) will work on one type of article, and people with different knowledge (e.g., healthcare policy) will likely work on other types. When these different types of contributors find themselves working on the same articles (i.e., the U.S. Affordable Care Act), their external contribution patterns reflect their diverse knowledge. It may also capture commonalities between contributors that go beyond subject matter expertise, such as a set of skills a contributor possesses or certain roles a contributor prefers to play (e.g., Liu and Ram 2011). For instance, contributors who prefer to contribute new content may work on different articles than those who like to edit or "shape" content for easier consumption, which may still differ from those who like to defend established content from unhelpful changes (Kane et al. 2014b).

The affiliation network approach accounts for both of these ways of understanding network value in a single approach (Faust 1997). Degree centrality in the affiliation network measures both the diversity of knowledge that can be transferred to the article and the diversity of the underlying knowledge and skills possessed by the contributors. We also weight the degree centrality by the quality of the article. This weighting accounts for the fact that contributors would seek to transfer better knowledge, not just more of it. Contributors who work on other high-quality articles also likely have more valuable skills for generating high-quality content than those that work on other low-quality ones. Thus, we hypothesize the following:

Hypothesis 2. *The quality-weighted degree centrality of an article in the affiliation network is positively associated information quality.*

Yet, different articles may be a more or less valuable source of knowledge. Contributors are more likely to contribute to heavily viewed articles (Zhang and Zhu 2011), leading to disproportionately high collaboration in a relatively small number of articles. Consistent with power-law distributions observed in other types of networks (Faraj and Johnson 2011, Johnson et al. 2014), the vast majority of articles on Wikipedia are relatively modest efforts involving only a few contributors, whereas others are robust collaborative efforts involving thousands of contributors (Capocci et al. 2006). The latter articles are likely to provide deeper resources for the peer production community, and the depth of resources available in a network is associated with the quality of information produced (Constant et al. 1996).

The *eigenvector centrality* in the affiliation network can measure the depth of resources available to the peer production community. Like our degree centrality measure, eigenvector centrality reflects both interpretations of the affiliation network (Faust 1997, Podolny and Baron 1997). First, it accounts for the level of collaborative activity occurring in the articles its contributors work on, resulting in higher centrality scores for articles whose contributors work on more active articles. Articles with high levels of contributor activity contain more robust resources that can be transferred to facilitate production of other articles. Second, eigenvector centrality reflects the reputation of highly active contributors in the affiliation network (Faust 1997), a reputation that may allow them to influence production more than others. Contributors working deeply on articles develop greater experience and skill that aid their work on other articles, allowing them to help develop higher-quality articles. Combining these effects indicates that deep contributors on active articles have the greatest effect on eigenvector centrality in the affiliation network.

Eigenvector centrality also reflects the position of the article with respect to all activity occurring in the affiliation network, not just in the articles it is directly connected to. Since the depth of collaboration occurring in one article can affect the other articles those contributors work on, the associated improvement in those other articles then cascades to other articles across the network. Indeed, network effects are often discernable multiple degrees of separation from a focal effect. For example, a strong recommendation for a particular product in a product recommendation network influences the value of the other products up to three degrees of separation (Oestreicher-Singer and Sundararajan 2012). Eigenvector centrality measures the relative position of an article to the collaborative activity occurring in the entire affiliation network, reflecting network effects at multiple degrees of separation. For this reason, some have called eigenvector centrality in an affiliation network "total effects centrality" (Friedkin 1991, Faust 1997). Thus, we hypothesize the following:

Hypothesis 3. *The eigenvector centrality of an article in an affiliation network is positively associated with the quality of information in the artifact.*

Research Method and Setting

Established in 2001, the English version of Wikipedia offers approximately 4.5 million distinct articles, as well as tens of millions of articles in hundreds of other languages. We focus our empirical analysis on Medicine WikiProject in Wikipedia, a relatively insular subcommunity focused on developing health and medical information. A WikiProject refers to a group of contributors who are dedicated to developing, maintaining, and organizing articles related to a particular topic. The Medicine WikiProject consists of 16,244 articles created by 147,362 unique contributors, during the 10-year period we studied between 2001 and 2011.

We focus on a single WikiProject for several reasons. First, studying a single WikiProject reduces the computational requirements for analysis. Conducting our longitudinal network analysis on the entire English Wikipedia—consisting of 4.5 million articles developed by 24.5 million nonanonymous contributors—is computationally prohibitive. Second, a random sample of articles would not yield the social network features of theoretical interest. A WikiProject provides clearly defined boundaries and norms for the network, which is an important consideration in SNA (Scott 2012). Third, WikiProjects usually rate the quality of each article under their purview, so we have access to a measure of relative information quality. However, quality ratings have different meanings for different WikiProjects, so they are difficult to compare across projects. Fourth, WikiProjects provide some experimental control for the subject matter. Articles related to medicine may be evaluated according to entirely different quality standards than articles on pop culture; studying articles dedicated to a particular WikiProject thus limits the impact of potentially confounding factors.

Dependent Variable

An additional advantage of choosing the Medicine WikiProject is the clear domain in which to externally validate the quality of information contained in the artifacts. Medical professionals and researchers have established standards by which to validate the quality of medical information (Eysenbach et al. 2002). The Medicine WikiProject provides an assessment of the quality of the information they produce. For the period of our analysis, the Medicine WikiProject evaluates articles on a seven-point scale (from lowest to highest quality: Stub, Start, C, B, Good, A, Featured). See Appendix B for a detailed description of these quality ratings and how they are selected. Although previous research shows that this rating system is a good proxy for information quality in other WikiProjects (Kittur et al. 2008), these studies have often relied on crowd-sourced methods to validate the quality of information. Although amateur crowds are often good at identifying inaccuracies in information, they are often not good at identifying when information is incomplete, often omitting critical information (Howe 2008, Clauson et al. 2008). Thus, expert evaluation is critical for validating the accuracy and completeness of medical information, critically important aspects of medical information quality (Eysenbach et al. 2002).

We also used a team of fourth-year medical students to validate the reliability of these quality ratings with respect to the objective quality of the underlying medical information. The medical students received a description of each quality designation and its associated quality standards. They evaluated paper printouts of the article content, divorcing content from its editorial history. In a small pilot evaluation with 14 articles, randomly selected from across each of the quality ratings, the medical students independently evaluated information quality on the same seven-point scale used by the Medicine WikiProject. The medical school students then compared their own results with each other, discussing the reasons for their evaluations as a team. In general, the students evaluated articles uniformly; in the rare cases of discrepancies, students discussed their reasons for applying their rating and ultimately reached consensus. We then compared the students' independent assessment with those applied by the WikiProject. In general, the ratings applied by the WikiProject matched those by the medical school students (see Appendix C for a further discussion).

After the initial pilot test, the medical school students evaluated a stratified sample of an additional 95 randomly selected medical articles, using the same method. They first developed their quality ratings independently (Pearson's $r = 0.78$). Then the students reconciled any disagreements and determined a single consensus rating for each article. We compared these ratings with those assigned by the WikiProject Medicine (Pearson's $r = 0.91$). The most frequent discrepancies involved distinguishing between start and stub quality articles, which suggests marginal differences between these two quality groupings. Thus, the ratings by the Medicine WikiProject appear to provide a reliable measure of relative information quality.

Independent Variables

We use three measures to assess collaboration—(a) the number of contributors, (b) the quality weighted degree centrality, and (c) the eigenvector centrality. First, the number of distinct contributors working on an article captures direct contributions to an information artifact, independent of people's contributions to other artifacts in the system. We measure this as the number of total contributors to an article for the past three months.

Second, we weight each edge in the network with the quality of the article at the time the contributor edited it and calculate the *degree centrality* of each article in the two-mode network. Since our theoretical argument is that better information flows through the network, it makes sense to place a greater emphasis on articles that are of higher quality. We weighted the connection to other articles made by a shared contributor by converting the quality rating of the connected article to an ordinal scale. A limitation of this aggregation is that it treats the intervals between ordinal ratings as the same. Although methods exist to handle specific cases of ordinal aggregation (e.g., Cook and Kress 1985), recent studies indicate that many conclusions are robust to parametric assumptions (Carifio and Perla 2008, Norman 2010). Similar sorts of ordinal weighting are common in SNA (Scott 2012). We further divide the weighted degree centrality by the number of distinct contributors to better isolate the diversity and strength from just the volume of contribution.

Third, *eigenvector centrality* captures the depth of information sources available, in that it summarizes the node's centrality among all nodes and ties that constitute the network. Since *eigenvector centrality* captures the relationship of each node to all other nodes in the network, it is particularly computationally intensive. Eigenvector centrality is defined as follows (Bonacich 2007):

> The centrality of a vertex i is proportional to the sum of the centralities of the vertices to which it is connected. λ is the largest eigenvalue of A and n is the number of vertices:

$$Ax = \lambda x, \qquad \lambda x_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, n.$$

To increase the likelihood that the empirical relationships we observe are causal, we measure the focal network variables for three months prior to the month that the dependent variable is measured. For example, we measure the number of distinct contributors prior to the measurement of article quality. Therefore, we reduce the likelihood that our models reflect a reverse causality from our hypotheses.

Control Variables

Because research in the content stream of information quality shows that factors other than the affiliation network affect article quality rating (e.g., Warncke-Wang et al. 2013), we control for the explanations of article quality used in this literature: article age, number of article views, length, reading complexity, section depth,

Table 1 Descriptive Statistics and Variable Correlations

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11
1. *Age* (ln, days)	6.91	1.01	1.00										
2. *Length* (ln, chars)	2.21	1.40	0.52	1.00									
3. *Complexity* (ARI)	19.89	5.47	0.07	0.28	1.00								
4. *Section depth*[a, b]	1.68	25.91	−0.02	−0.28	−0.17	1.00							
5. *External references*[a, b]	0.75	0.88	0.26	0.44	0.15	−0.05	1.00						
6. *Internal links*[a, b]	7.20	5.49	−0.17	−0.63	−0.14	0.09	−0.32	1.00					
7. *Multimedia content*[a, b]	3.10	31.71	0.01	−0.01	−0.02	0.00	−0.01	0.02	1.00				
8. *Anonymity* (%)	0.30	0.16	0.44	0.25	0.03	−0.01	−0.07	−0.10	0.00	1.00			
9. *Distinct contributors*	92.82	190.48	0.39	0.38	0.07	0.00	0.21	−0.07	−0.01	0.38	1.00		
10. *Degree centrality*[b]	46.09	251.07	−0.17	0.05	0.02	−0.01	0.09	−0.05	0.00	−0.15	−0.06	1.00	
11. *Eigenvector centrality*	0.04	1.42	0.10	0.26	0.06	−0.01	0.08	−0.05	−0.01	0.15	0.47	0.19	1.00
12. *Quality rating*	2.87	0.74	0.00	0.13	0.03	0.00	0.01	0.00	−0.02	0.01	0.10	0.03	0.18

Notes. Correlations refer to 161,800 monthly observations of 2,475 Wikipedia Medicine articles observed from February 2001 until August 2011.
[a]Indicates the measures is divided by the article length.
[b]Indicates the variable is scaled for presentation.

number of external references, amount of multimedia content, and anonymity of contributors. Table 1 provides a descriptive overview of these data. These controls help to distinguish whether communities are creating higher quality information, not just more of it.

Because article maturity likely relates to information quality, we control for article *age*. Articles in existence for a longer time have had a greater opportunity to support collaboration and attract additional contributors. For our sample, article ages range from one day to 10.4 years, with an average of 3.8 years. The *length* of the article could relate to information quality too. An information artifact thus may be perceived as higher in quality, simply because it has more, not just better, information. To control for this explanation, we control for the length of each article, expressed in thousands of characters of text, such that the measure ranged from 0 (Stub articles) to 1,094,011 characters. We use the natural log of article length in the statistical models.

Perceptions of higher quality also might derive from a more sophisticated writing style. That is, articles may be perceived as containing high-quality information if they sound authoritative, whether or not they actually are. Alternatively, articles could be incomprehensible if they are excessively difficult to read. To control for possible stylistic differences, we control for reading complexity, which we measured using the Automated Readability Index (ARI). (As a robustness check, we also compared the models using the Coleman-Liau index but found no substantive changes.) The ARI is calculated as

$$ARI = \left(\frac{4.71 \times letters}{words} + \frac{0.5 \times words}{sentences} - 21.43 \right).$$

The formula estimates the U.S. school grade reading level required to understand a sample of text. In our analysis, the relative level of this measure is used to differentiate simpler text from more complex text. For presentation purposes, we divide the reading complexity score by 1,000.

People can contribute to articles regardless of whether they log in to the Wikipedia system. We control for the influence of anonymity in a collaborative environment. When a contributor changes an article, her identity is automatically recorded in the system. If the contributor is not logged in to the system, this identity is recorded as an anonymous IP address. Because the raw number of anonymous contributors was too highly correlated with the total number of contributors, we use the *percentage of anonymous contributors*, calculated as the total number of anonymous contributors divided by the total number of contributors to an article. On average, anonymous contributors offered 30.3% of the contributions per article.

Wikipedia policy states that all information contributed must be supported by an authoritative external reference. Elsewhere on Wikipedia, authoritative references typically include any reputable published source (e.g., books, mainstream news media), but on the WikiProject Medicine, only peer-reviewed medical journals are considered authoritative. On average, articles in our sample included 19.3 external references. Similarly, articles can also link directly to other articles within Wikipedia; we control for these through a measure of *internal links*. In our sample, articles linked to an average of 89.1 other articles. Both link measures tend to increase with article length; therefore, we divide each by the article length so that external and internal length density is measured.

Finally, different forms of information presentation may influence the perception of quality, without actually influencing the quality of the information itself. Some articles appear to contain more or less valuable information because they contain images, sounds, or other multimedia information. To control for this effect, we measure the total number of *multimedia* images in the article. Similarly, the organization of an article may influence perceived quality, such that well-organized information is more accessible to the reader. Articles in Wikipedia can contain up to six levels of nested sections. We therefore include the maximum section level reached in an article and refer to this measure as *section depth*. Again, because both measures tend to increase with article length, we divide by the article length.

Results

For the full empirical analysis, we downloaded the full text history of 2,677,397 revisions of 16,244 articles in the Medicine WikiProject from its beginning (February 2001) until August 2011, which provided a 104.3 GB set of raw data. We employed a 114-node Linux cluster to allow for simultaneous downloading and processing of these extensive data. For each contribution to an article, the contributor's identity, changes made to the article, a description of the change, and the time of the change were all recorded. Within this sample, automated background processes instituted 2% of the changes (43,309 revisions), which we excluded from our analysis. (These automated processes perform routine mundane tasks such as checks for copyright violation, corrections of numbering, or searches for vandalism.) We took advantage of the remaining full revision history to construct the two-mode affiliation network, which linked articles through their contributors.

Similar to other studies of online peer production communities (e.g., Grewal et al. 2006, Trier 2008, Singh and Tan 2011), we first visualize the network relationships, because the network may exhibit structures in the social graph that would be overlooked in a purely statistical analysis. The size of our data set of 16,244 articles and 147,362 distinct contributors was intractable for most common visualization programs and would be difficult to interpret, so we created a small subsample of articles to create the visualization. Specifically, we took all Featured, A-quality, and Good articles (approximately 100) and paired them with a matched sample of B-quality and Start-quality articles. We obtained the identity and number of contributions for all contributors who had made at least four changes to any of these articles, reducing the population from the original 1,800 contributors in our data. These data enabled us to construct an edgelist, which we imported into UCINet 6.198 to construct the two-mode matrix of articles and contributors. Appendix A offers a visual representation of the resulting two-mode network, constructed using NetDraw 2.081 through an iterative process that positioned each node in the network according to its relationships to all other nodes, using a multidimensional scaling algorithm (Huisman and Van Duijn 2005). This visualization showed a striking relationship between the centrality of an artifact in the network and its quality, as evidenced by the cluster of high-quality articles in the center of the network.

Figure A.1 in Appendix A also shows a cluster of moderate quality articles surrounding a single central editor. Further research found that this contributor is a Dutch doctor working in the United Kingdom, who is committed to developing quality medical information on Wikipedia. We verified the authenticity of this profile by corroborating it with active profiles on other social media sites (e.g., LinkedIn and Twitter), official websites (e.g., the UK National Health Service), and published research using the purported author's name. This verification process gives confidence that this Wikipedia contributor was who he claimed to be offline. An examination of this contributor's activity suggests that he starts numerous articles on missing topics, develops them to acceptable levels of basic quality, and then starts new articles. He has worked on a greater number of articles than any other member of the Medicine WikiProject and appears to be employing a satisficing strategy, contributing a high volume of "good enough" information to increase the volume of quality information found in the WikiProject. He also may be seeding content for others to work on. Thus, although this visualization of our data suggests initial support for our hypotheses, our visualization also indicates that there are interesting aspects of information quality in online peer production settings that may be overlooked by analysis of the complete affiliation network.

We then analyzed the full data set using *R*. Table 2 describes the full results of a Bayesian ordinal logistic regression for article quality. Ordinal regression is appropriate when there is a progressive relationship within a categorical dependent variable but the magnitude of difference between categories is uncertain. For example, we might know which Olympics athletes won gold, silver, or bronze medals without knowing the final scores of any of the athletes. A large number of the articles never progressed beyond the Stub or Start quality. We exclude these articles with limited or no content or collaboration; including them would both artificially inflate the predictive power of our models and obscure nonobvious results. Therefore, although we compute network measures on the entire WikiProject Medicine project, we focus our empirical analysis on 161,800 monthly observations of the 2,475 articles that progress beyond Stub and Start from February 2001 until August 2011. This approach also raises the difficulty on our analysis method but further ensures that our affiliation network analysis is actually providing insight regarding information quality, not merely separating the best articles from article fragments. We assess support for our hypotheses by evaluating the confidence intervals around the

Table 2 Quality of Articles

Variable	Model 0	Model 1	Model 2	Model 3	Model 4
Age (ln, days)	−0.390***	−0.429***	−0.357***	−0.345***	−0.356***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Length (ln, characters)	0.889***	0.774***	0.857***	0.750***	0.688***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
Complexity (ARI)	−0.103***	−0.090***	−0.097***	−0.091***	−0.082***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Section depth	0.176***	0.150***	0.170***	0.149***	0.135***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
External references	−0.241***	−0.268***	−0.247***	−0.239***	−0.257***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Internal links	0.465***	0.401***	0.456***	0.399***	0.371***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Multimedia content	−0.043***	−0.039***	−0.043***	−0.040***	−0.038***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Anonymity (percentage)	0.600***	0.050	0.663***	0.419***	0.165***
	(0.038)	(0.041)	(0.038)	(0.038)	(0.041)
Distinct contributors		0.585***			0.401***
		(0.016)			(0.016)
Distinct contributors (squared)		−0.430***			−0.363***
		(0.013)			(0.014)
Degree centrality			0.152***		0.100***
			(0.008)		(0.007)
Eigenvector centrality				0.303***	0.258***
				(0.006)	(0.007)
Residual deviance	249,086.502	247,653.495	248,548.872	246,520.715	245,620.471
AIC	249,110.502	247,681.495	248,574.872	246,546.715	245,652.471
BIC	249,230.432	247,821.413	248,704.796	246,676.638	245,812.377

Notes. Bayesian ordinal logit regressions of article quality; 161,800 monthly observations of 2,475 articles from February 2001 until August 2011; standard errors in parentheses; 1,000 iterations. All variables are standardized, except Anonymity; asterisks indicate that zero lies outside a 99.9%(***) confidence interval.

estimated model coefficients. Model 0 contains only the control variables.

Model 1 indicates support for our first hypothesis, in that network resources have a curvilinear effect on artifact quality. Both the linear and squared coefficients are significant ($\beta = 0.585$ and $\beta = -0.430$, respectively). These coefficients thus indicate an inverted U-shaped relationship with article quality.[1] We find that additional contributors working on an information artifact increase quality up to an optimal point, but then more contributors detract from the quality of information in the artifact. In Model 2, we find support for our second hypothesis pertaining to how the diversity of resources increases artifact quality. The coefficient for quality-weighted degree centrality is positive and significant ($\beta = 0.152$), such that more diverse knowledge accessed or possessed by contributors tends to improve the quality of the information artifact. This result shows that the number of other articles contributors work on influences the quality of their collaborative activity. Model 3 shows that the coefficient on eigenvector centrality is positive and significant ($\beta = 0.303$), supporting Hypothesis 3. Greater depth of resources available through an artifact or represented by its contributors leads to a likely higher quality of the artifact. As a global measure of centrality in the affiliation matrix, this measure suggests that resources available in the entire affiliation network affect article quality—not just those to which an article or its contributors are directly connected.

Model 4 contains the results with all focal and control variables, and it continues to offer support for our hypotheses. Examination of variance inflation factors does not indicate potential for multicollinerarity; the largest factor is 3.3. The magnitude of the coefficients indicates that the coefficients of our focal measures are of the same approximate magnitude as the control variables, so our focal measures are not only statistically

[1] We also tested different specifications of this variable, finding that our curvilinear formulation best fits the data. The curvilinear model improves the model deviance by 701.18 over an unreported model that includes only a linear term and by 22.57 over an unreported model that includes a log-linear term.

but also practically significant. Because the focal network metrics are standardized, they can be interpreted relative to their distribution in our sample. In Model 4, we see that a one standard deviation increase in quality weighted degree centrality is associated with a 10% ($e^{0.100} - 1$) increase in likelihood of a higher quality rating. Similarly, a one standard deviation increase in eigenvector centrality is associated with a 29% ($e^{0.258} - 1$) increase in likelihood of a higher quality rating.

As a robustness check, we validate these results using latent state models (Appendix D), which provide qualitatively similar results.

Alternative Quality Assessment

Although the assessment of a small number of articles by medical students was effective, it would not scale to handle the volume of articles in Wikipedia. Similarly, the ratings done by contributors to the Medicine WikiProject requires expertise. Crowdsourcing methods of assessing article quality have been effective for numerous tasks, including Wikipedia article rating (Kittur et al. 2008). However, medical articles—with clear scientific basis and verifiable claims—provide an opportunity to investigate more deeply into the nature of these quality ratings.

To investigate whether crowdsourcing would be a reliable metric for medical articles, we asked Mechanical Turk (MTurk) to assign four workers to evaluate the quality of each article that the medical students evaluated. We used the same protocol as Experiment 2 employed by Kittur et al. (2008) to assess the quality of the articles. Workers completed a set of four factual questions about the article (to assess attention to the task), rated the article on a seven-point scale (worded closely to the WikiProject medicine criteria), and gave a brief justification for the rating. We obtained 386 usable ratings that demonstrated attention to the task; each article was rated by at least three workers. Table 3 describes the results of a Bayesian ordinal regression on article, the same analysis method we used in our base models.

Model 0 includes only the MTurk rating and finds positive and significant ($\beta = 0.351$) explanatory value from the rating, seeming to indicate that MTurk workers are effective at assessing medical article quality. Additional analysis, however, begins to shed doubt on that interpretation. Model 1 adds variables that measure article content, and the MTurk assessment of quality is no longer significant when controlling for these content characteristics. These results indicate that, for medical articles, MTurk workers may be making their quality ratings primarily by observing easily visible content measures. Furthermore, since the Akaike Information Criterion (AIC) and Bayesian Information

Table 3 Mechanical Turk Prediction

Variable	Model 0	Model 1	Model 2	Model 3
MTurk Rating	0.351***	0.071	−0.080	0.259*
	(0.068)	(0.101)	(0.141)	(0.106)
Age (ln, days)		0.089		
		(0.180)		
Length (ln, characters)		3.125***		
		(0.347)		
Complexity (ARI)		−0.785**		
		(0.270)		
Section depth		−0.420		
		(0.729)		
External references		1.169***		
		(0.197)		
Internal links		1.529***		
		(0.310)		
Multimedia content		−3.690***		
		(0.809)		
Anonymity (percentage)		0.779		
		(1.081)		
Observations	386	386	151	151
Residual deviance	696.273	411.057	207.457	280.011
AIC	706.273	437.057	215.457	290.011
BIC	726.052	488.483	227.526	305.097

Notes. Bayesian ordinal logit regressions of article quality; standard errors in parentheses; 1,000 iterations. All continuous variables are standardized, except Anonymity; asterisks indicate that zero lies outside a 95%(*), 99%(**), or 99.9%(***) confidence interval.

Criterion (BIC) measures are significantly worse in models using MTurk, it indicates that easily observable content characteristics may actually be a more effective measure of quality.

Furthermore, Models 2 and 3 provide additional insight. In Model 3, we remove all of the articles that are rated Start or Stub, and the MTurk assessments of quality are no longer significant. The significant effect of the MTurk workers comes from distinguishing the best and the worst articles from one another, but not from finer grained distinctions between complete articles. To rule out the possibility that this lack of significance stems from the reduced sample size of the higher quality articles, we use a random sample of articles in Model 3 that matches the sample size in Model 2. The MTurk assessment of articles is again significant. The results in Table 3 suggest that the MTurk assessment of article quality is mainly a signal of the underlying content characteristics of the article, distinguishing between early and complete articles. These findings add to the ongoing literature regarding the value of Mechanical Turk ratings (e.g., Paolacci et al. 2010, Buhrmester et al. 2011).

Discussion

This paper argues that an affiliation network is a productive way to assess information quality on Wikipedia, because it can overcome drawbacks associated with other major streams of assessing information quality pursued independently. We treat the 16,244 medical articles on Wikipedia as nodes in a network, created by the 147,362 contributors who have worked on one or more of the articles. We find support for our hypotheses. Specifically, the total number of contributors is curvilinearly related to information quality, such that more contributors are better but only up to a point. The quality-weighted degree centrality of the article in the affiliation network, capturing the diversity of knowledge present in the peer production community by assessing the other articles contributors work on, is also positively related to information quality. Last, the eigenvector centrality of the article in the affiliation network, capturing the depth of knowledge in the network that is available to the peer production community, is significant as well. The resulting picture is one of multiple collaborators working on multiple projects in a peer production ecosystem, in which collaboration in one article affects the information quality in another.

Theoretical Implications

This paper has a number of theoretical implications. First, it does find that affiliation networks that treat both contributors and knowledge artifacts as part of a single network are a productive conceptualization for studying knowledge creation in online peer production communities. This approach has a number of benefits that build on previous approaches to studying information quality. Specifically, it accounts for content as a part of the network, recognizing that the characteristics of content may actually influence the type of collaboration that takes place in a particular peer production community. It also recognizes that contributors to many types of online peer production communities frequently participate in multiple communities simultaneously, and the activity in one community may influence production in another. Thus, the affiliation network serves as a succinct approach to assessing information quality that accounts for multiple mechanisms simultaneously.

Second, our study also finds that information quality is a separate characteristic from simple content measures. By controlling for the content characteristics in our models, we separate information quality from these readily observable characteristics, such as number of words or external references. Furthermore, by validating the information quality metrics using external coders, we have greater confidence that we are actually establishing a connection between affiliation network structure and information quality. We also find that some alternate measures of information quality, such as MTurk ratings, serve as a proxy for content characteristics and not independent measures of quality. Future research using MTurk ratings should guard against such a possibility in the future.

Another implication is that researchers should regard these types of online peer production communities as an ecosystem of contributors and content, rather than a series of independent freestanding communities. How the communities and groups relate to each other may be a factor in their successful production. Researchers, therefore, may start paying attention to how these groups work together as an integrated network. This rationale may suggest cultivating different types of online production communities specifically for the purpose of developing cross-community memberships that can facilitate production. For example, many companies establish extra-professional groups in their online collaboration systems dedicated to certain hobbies like cycling or cooking. Our results suggest that if these extra-professional communities attract different types of members than work related or project groups, part of their value may be in creating connections that allow information to flow in different ways through the organization and be made available to work-related groups that may need it.

Limitations and Future Research

This study also has a number of limitations that can yield future research. First, although we believe the ability of the affiliation network to account for two different mechanisms of how networks affect information quality, it nevertheless cannot distinguish between these two possible mechanisms. We do not disaggregate these mechanisms in this study, as theoretical reasons suggest that both mechanisms influence the quality of information developed by the community, which is our research question. Additional research would be needed, however, if researchers desire a more detailed understanding of the relative effect of each mechanism.

Second, our study focuses on medical articles developed by a particular WikiProject dedicated to this topic. Although we believe that the overall affiliation network effects to be generalizable to other communities, some of the more nuanced findings about information quality and MTurk may be particular to medical articles or to this WikiProject. We have theoretical reasons for drawing the network boundaries at the WikiProject level and such boundaries allow other types of analysis (i.e., quality validation by experts). Only a longitudinal network analysis of all of Wikipedia could entirely rule out these alternative explanations, analysis that is too computationally intensive for the resources we had at our disposal.

Third, although we find the affiliation network to be a powerful tool for assessing information quality on Wikipedia, it clearly does not tell the entire story of information quality on Wikipedia. Our multiple methods of analysis (e.g., qualitative visualization) revealed several interesting aspects of collaboration that were overlooked by our large-scale quantitative methods. Future research should augment these quantitative analyses with deeper qualitative study (e.g., Kane 2011, Kane et al. 2014b) to obtain the clearest picture of collaboration and information quality in peer production communities.

Conclusion

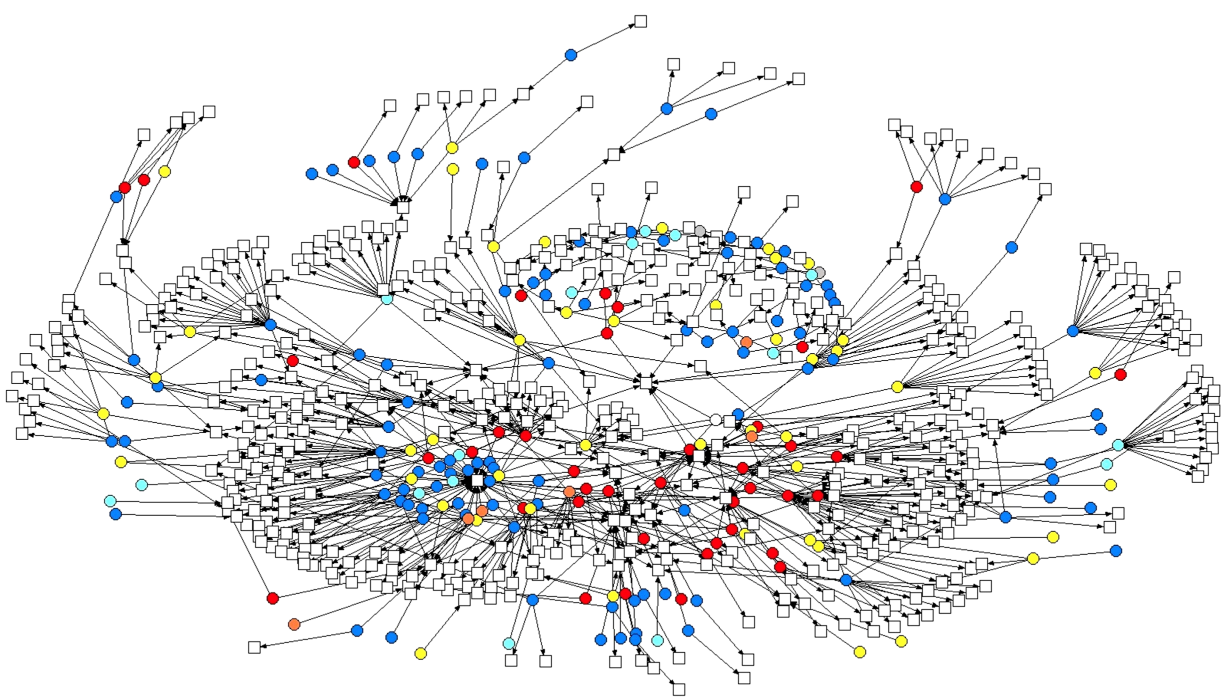
By adopting a social network perspective, this research demonstrates that collaboration that occurs across information artifacts may be critical to the interactions that occur on online peer production platforms. Researchers should consider the social structure that defines all elements of a collaborative platform, rather than simply those elements directly related to the particular outcome of interest. Managers in turn need to manage their collaborative platforms as single, integrated environments, rather than as portfolios of independent collaborative efforts.

Acknowledgments

The first author gratefully acknowledges funding support for this research from the National Science Foundation [CAREER Award 0953285].

Appendix A.

Figure A.1 (Color online) Two-mode Network of Articles and Contributors



Squares = Contributors
Circles = Articles
Red = Featured articles
Orange = A-quality articles
Yellow = Good articles
Dark blue = B-quality articles
Light blue = Start-quality articles

Appendix B. Quality Ratings Used by Medicine WikiProject

Rating	Criteria	Reader's experience	Editing suggestions	Example
Featured article	The article has attained featured article status by passing an official review.	Professional, outstanding, and thorough; a definitive source for encyclopedic information.	No further content additions should be necessary unless new information becomes available; further improvements to the prose quality are often possible.	Tourette syndrome
A-Class	The article is well organized and essentially complete, having been reviewed by impartial reviewers from this WikiProject or elsewhere. Good article status is not a requirement for A-Class.	Very useful to readers. A fairly complete treatment of the subject. A nonexpert in the subject would typically find nothing wanting.	Expert knowledge may be needed to tweak the article, and style problems may need solving. Peer review may help.	Not currently used by this project (*used though* 2011)
Good article	The article has attained good article status by passing an official review.	Useful to nearly all readers, with no obvious problems; approaching (but not equaling) the quality of a professional encyclopedia.	Some editing by subject and style experts is helpful; comparison with an existing featured article on a similar topic may highlight areas where content is weak or missing.	Common cold
B-Class	The article is mostly complete and without major problems, but requires some further work to reach good article standards.	Readers are not left wanting, although the content may not be complete enough to satisfy a serious student or researcher.	A few aspects of content and style need to be addressed. Expert knowledge may be needed. The inclusion of supporting materials should also be considered if practical, and the article checked for general compliance with the Manual of Style and related style guidelines.	Tuberous sclerosis
C-Class	The article is substantial, but is still missing important content or contains much irrelevant material. The article should have some references to reliable sources, but may still have significant problems or require substantial cleanup.	Useful to a casual reader, but would not provide a complete picture for even a moderately detailed study.	Considerable editing is needed to close gaps in content and solve cleanup problems.	Islets of Langerhans
Start	An article that is developing, but which is quite incomplete. It might or might not cite adequate reliable sources.	Provides some meaningful content, but most readers will need more.	Providing references to reliable sources should come first; the article also needs substantial improvement in content and organization. Also improve the grammar, spelling, writing style, and improve the jargon use.	AIDS-related lymphoma
Stub	A very basic description of the topic. However, all very-bad-quality articles will fall into this category.	Provides very little meaningful content; may be little more than a dictionary definition. Readers probably see insufficiently developed features of the topic and may not see how the features of the topic are significant.	Any editing or additional material can be helpful. The provision of meaningful content should be a priority. The best solution for a Stub-class article to step up to a Start-class article is to add in referenced reasons of why the topic is significant.	Acute care

Source. Reproduced from http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Medicine/Assessment.

Appendix C. Quality Ratings on Wikipedia

Wikipedia assigns two ratings, Featured and Good, through a peer review process. This process begins when someone nominates an article for inclusion for consideration as a Featured or Good article. Then an open peer review process begins, reminiscent of an academic peer review, to determine whether the article is of sufficiently high quality to be granted Featured/Good article status. They may vote to support or oppose the nomination and must cite the reasons for their vote. An elected committee of Wikipedia administrators oversees and moderates this debate. Only when all issues are successfully resolved is the article promoted by consensus. Of the 4.5 million Wikipedia articles in English, 0.09% (approximately 4,503) have achieved Featured article status,[2] and 0.44% (approximately 21,878) have received a Good article rating.[3] Members of the Medicine WikiProject assign the other ratings (A, B, C, Stub, and Start) by consensus. Some A-level articles also received Good article ratings (A-level is a higher quality rating than Good), but no other ratings

[2] http://en.wikipedia.org/wiki/Wikipedia:Featured_articles.

[3] http://en.wikipedia.org/wiki/Wikipedia:Good_articles.

received any formal quality recognition by Wikipedia at large. The WikiProject has an assessment committee who manages the quality ratings for the articles in the WikiProject.

In general, the ratings applied by the WikiProject matched those by the medical school students, with one exception: the students believed that the ratings applied to C-quality articles were not reliable. It had been relatively recently introduced to the rating system (in 2008), and C-quality referred to full-length articles with some flaws, usually in that they contained extraneous information unrelated to the article topic. The medical school students described the C-quality rating as a nonordinal "catch-all" rating for articles that did not fit neatly into another quality rating. They noted that the C-quality articles exhibited a wide range of article characteristics, whose quality was not easily identifiable. To confirm this perception, we asked the students to review five additional C-quality articles. This follow-up analysis revealed that C-quality articles did not provide a reliable proxy for information quality and were not appropriate for ordinal evaluation. Given this evaluation, we decided to exclude these articles from further quality validation and statistical analysis.

As a robustness check, we tested the final statistical model both with and without C-quality articles and found no meaningful difference in the results. Thus, our independent validation of information quality suggests that, although the methods the Medicine WikiProject used for evaluating information quality are generally reliable, validation by outside expert reviewers provide insight as to how these evaluations are applied in practice and what they mean for the underlying quality of information contained in them. Although this insight did not ultimately affect the broad results of our analysis, it is nevertheless a notable insight regarding information quality in online peer production settings, particularly if patients are trying to assess whether the medical information in one of the 3,598 articles with this rating should be relied on for health decisions.

Appendix D. Latent State Models

It is still possible that although information quality may be reflected in a Wikipedia rating, it may be imperfectly so. A shared underlying characteristic could bias both Wikipedia editors and our expert codes. To help rule out this explanation, we also test the robustness of our models by modeling information quality as unobservable latent characteristics, reflected in multiple observable factors. Hidden Markov models offer an alternative for assessing the effects of collaboration activity on information quality, allowing us to model information quality as a unobservable, latent characteristic that is manifest in multiple observable characteristics of the article (Singh et al. 2011b). For each article, we employ two repeated (monthly) measures of information quality. First, we use the ordinal rating given by the WikiProject. Second, we measure the number of page views of the article as another indicator of quality, since it is likely that readers would be attracted to higher quality articles (Ransbotham et al. 2012). The hidden Markov models infer the latent quality from these two signals and, importantly, measure the effects of covariates as articles move to higher quality states.

Finally, a transition model (φ) reflects the probability (φ_{ij}) of transition from state S_i in time t to state S_j in time $t+1$.

Table D.1 **Hidden Markov Model Analysis of the Effect of Networks on Quality Transition**

	Model 0		Model 1	
Variable	State S1	State S2	State S1	State S2
	Initial states			
Probability	0.483	0.517	0.490	0.510
	Unadjusted transition matrix			
State 1 (S1)	0.999	0.001	1.000	0.000
State 2 (S2)	0.001	0.999	0.001	0.999
	Effect of covariates on transition probabilities			
Age (ln, days)	−4.408	0.470	−0.614	6.917
Length (ln, characters)	0.295	−1.053	1.114	0.475
Complexity (ARI)	−0.144	0.124	0.045	−0.104
Section depth	4.248	−0.214	0.230	−2.940
External references	0.826	−0.239	0.164	−0.492
Internal links	−0.688	0.788	−0.947	0.968
Multimedia content	5.998	0.672	−0.961	−0.079
Anonymity (percentage)	0.722	−0.754	0.908	1.911
Distinct contributors			1.939	−1.900
Distinct contributors (squared)			−3.528	−3.860
Degree centrality			0.045	0.083
Eigenvector centrality			−0.479	0.274
Log-likelihood		−286,012.9		−285,914.8
AIC		572,087.8		571,907.6
BIC		572,380.2		572,275.4
Log-likelihood ratio X^2				196.249***

Notes. Two-state hidden Markov model where S1 is a lower quality state and S2 is a higher quality state.
***$p < 0.001$.

For each source (i) and destination state (j), the transitions probabilities sum to 1; $\sum_{i \in S} \varphi_{ij} = 1, \forall j \in S$. The hidden Markov model allows the transition probabilities to depend on each of the available covariates. These transition probabilities are the focus of the empirical analysis and estimate the influence of individual covariates on the underlying information quality. Parameters for the model were estimated with an expectation maximization algorithm (Visser and Speekenbrink 2010). For tractability and ease of interpretation, Table D.1 focuses on a two-state (lower quality and higher quality) model. Table D.1 describes the effects of covariates on the transition from the lower quality state to a higher quality state.

The unadjusted transition matrix indicates that the states are quite sticky. This makes sense; changes in content or collaboration are usually required to change quality. Model 1 shows the effect of the focal collaboration covariates on the transitions between quality states. Although these models do not support estimates of statistical significance, the only difference between Model 0 and Model 1 is the inclusion of the collaboration covariates and the log-likelihood ratio test indicates significantly better fit ($\chi^2 = 196.249$, $p < 0.001$). So, the affiliation network measures improve the models by a statistically significant amount. Additionally, Model 1 column S2 indicates that the signs for the collaboration variables are the same as in the ordinal regression.

References

Arazy O, Nov O, Patterson R, Yeo L (2011) Information quality in Wikipedia: The effects of group composition and task conflict. *J. Management Inform. Systems* 27(4):71–98.

Asvanund A, Clay K, Krishnan R, Smith MD (2004) An empirical analysis of network externalities in peer-to-peer music-sharing networks. *Inform. Systems Res.* 15(2):155–174.

Blumenstock JE (2008) Size matters: Word count as a measure of quality on Wikipedia. *Proc.* 17*th Int. Conf. World Wide Web* (ACM, New York), 1095–1096.

Bonacich P (2007) Some unique properties of eigenvector centrality. *Soc. Networks* 29(4):555–564.

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Sci.* 323(5916):892–895.

Brandes U, Kenis P, Lerner J, Raaij D (2009) *Network Analysis of Collaboration Structure in Wikipedia* (ACM, New York), 731–740.

Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. *Comput. Networks ISDN Systems* 30(1–7):107–117.

Buhrmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspect. Psych. Sci.* 6(1):3–5.

Butler BS (2001) Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Inform. Systems Res.* 12(4):346–362.

Capocci A, Rao F, Caldarelli G (2008) Taxonomy and clustering in collaborative systems: The case of the on-line encyclopedia Wikipedia. *Europhysics Lett.* 81(2):Article 28006.

Capocci A, Servedio VDP, Colaiori F, Buriol LS, Donato D, Leonardi S, Caldarelli G (2006) Preferential attachment in the growth of social networks: The Internet encyclopedia Wikipedia. *Phys. Rev. E* 74(3):Article 036116.

Carifio J, Perla R (2008) Resolving the 50 year debate around using and misusing Likert scales. *Medical Ed.* 42(12):1150–1152.

Carrington PJ, Scott J, Wasserman S (2005) *Models and Methods in Social Network Analysis* (Cambridge University Press, Cambridge, UK).

Clauson KA, Polen HH, Boulos MK, Dzenowagis JH (2008) Scope, completeness, and accuracy of drug information in Wikipedia. *Ann. Pharmacotherapy* 42(12):1814–1821.

Constant D, Sproull L, Kiesler S (1996) The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organ. Sci.* 7(2):119–135.

Cook WD, Kress M (1985) Ordinal ranking with intensity of preference. *Management Sci.* 31(1):26–32.

Davis A, Gardner BB, Gardner MR (1941) *Deep South* (University of Chicago Press, Chicago).

Eysenbach G, Powell J, Kuss O, Sa E-R (2002) Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *J. Amer. Medical Assoc.* 287(20):2691–2700.

Faraj S, Johnson SL (2011) Network exchange patterns in online communities. *Organ. Sci.* 22(6):1464–1480.

Faraj S, Jarvenpaa SL, Majchrzak A (2011) Knowledge collaboration in online communities. *Organ. Sci.* 22(5):1224–1239.

Faust K (1997) Centrality in affiliation networks. *Soc. Networks* 19(2):157–191.

Fershtman C, Gandal N (2011) Direct and indirect knowledge spillovers: The "social network" of open source projects. *RAND J. Econom.* 42(1):70–91.

Friedkin NE (1991) Theoretical foundations for centrality measures. *Amer. J. Soc.* 96(6):1478–1504.

Friedkin NE (1982) Information flow through strong and weak ties in intraorganizational social networks. *Soc. Networks* 3(4):273–285.

Gallaugher J, Ransbotham S (2010) Social media and customer dialog management at Starbucks. *MIS Quart. Executive* 9(4):197–212.

Granovetter MS (1973) The strength of weak ties. *Amer. J. Soc.* 78(6):1360–1380.

Grewal R, Lilien GL, Mallapragada G (2006) Location, location, location: How network embeddedness affects project success in open source systems. *Management Sci.* 52(7):1043–1056.

Halatchliyski I, Moskaliuk J, Kimmerle J, Cress U (2010) Who integrates the networks of knowledge in Wikipedia? *Proc.* 6*th Int. Symp. Wikis Open Collaboration* (ACM, New York), 1.

Hansen MT, Haas MR (2001) Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Admin. Sci. Quart.* 46(1):1–28.

Hiltz SR, Turoff M (1985) Structuring computer-mediated communication systems to avoid information overload. *Comm. ACM* 28(7):680–689.

Howe J (2008) *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business* (Crown Business, New York).

Huisman M, Van Duijn MAJ (2005) Software for social network analysis. Carrington PJ, Scott J, Wasserman S, eds. *Models and Methods in Social Network Analysis* (Cambridge University Press, New York), 270–316.

Iba T, Nemoto K, Peters B, Gloor PA (2010) Analyzing the creative editing behavior of Wikipedia editors: Through dynamic social network analysis. *Procedia–Soc. Behavioral Sci.* 2(4):6441–6456.

Johnson SL, Faraj S, Kudaravalli S (2014) Emergence of power laws in online communities: The role of social mechanisms and preferential attachment. *MIS Quart.* 38(3):795–808.

Jurgens D, Lu T-C (2012) Temporal motifs reveal the dynamics of editor interactions in Wikipedia. *Proc. 6th Internat. AAAI Conf. Weblogs Social Media* (AAAI Press, Menlo Park, CA), 162–169.

Kane GC (2011) A multimethod study of information quality in wiki collaboration. *ACM Trans. Management Inform. Systems* 2(1):Article 4.

Kane GC, Fichman RG (2009) The shoemakers children: Using wikis for IS research, teaching, and publication. *MIS Quart.* 33(1):1–22.

Kane GC, Johnson J, Majchrzak A (2014b) Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. *Management Sci.* 60(12):3026–3048.

Kane GC, Alavi M, Labianca G, Borgatti S (2014a) What's different about social media networks? A framework and research agenda. *MIS Quart.* 38(1):275–304.

Keegan B, Gergle D, Contractor N (2013) Hot off the wiki: Structures and dynamics of Wikipedia's coverage of breaking news events. *Amer. Behavioral Scientist* 57(5):595–622.

Kittur A, Kraut RE (2008) *Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination* (ACM, New York).

Kittur A, Kraut RE (2010) Beyond Wikipedia: Coordination and conflict in online production groups. *Proc.* 2010 *ACM Conf. Computer Supported Cooperative Work* (ACM, New York), 215–224.

Kittur A, Chi EH, Suh B (2008) Crowdsourcing user studies with Mechanical Turk. *Proc. SIGCHI Conf. Human Factors Computing Systems* (ACM, New York), 453–456.

Kittur A, Lee B, Kraut RE (2009) Coordination in collective intelligence: The role of team structure and task interdependence. *CHI 2009* (ACM, New York), 1495–1504.

Kuk G (2006) Strategic interaction and knowledge sharing in the KDE developer mailing list. *Management Sci.* 52(7):1031–1042.

Labianca G, Brass DJ (2006) Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. *Acad. Management Rev.* 31(3):596–614.

Lipka N, Stein B (2010) Identifying featured articles in Wikipedia: Writing style matters. *Proc.* 19*th Int. Conf. World Wide Web* (ACM, New York), 1147–1148.

Liu J, Ram S (2011) Who does what: Collaboration patterns in the Wikipedia and their impact on article quality. *ACM Trans. Management Inform. Systems* 2(2):Article 11.

Maniu S, Cautis B, Abdessalem T (2011) Building a signed network from interactions in Wikipedia. *Databases Soc. Networks* (ACM, New York), 19–24.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

Massa P (2011) Social networks of Wikipedia. *Proc.* 22*nd ACM Conf. Hypertext Hypermedia* (ACM, New York), 221–230.

Nemoto K, Gloor P, Laubacher R (2011) Social capital increases efficiency of collaboration among Wikipedia editors. *Proc.* 22*nd ACM Conf. Hypertext Hypermedia* (ACM, New York), 231–240.

Norman G (2010) Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Ed.* 15(5):625–632.

O'Dell C, Grayson CJ (1998) If only we knew what we know: Identification and transfer of internal best practices. *California Management Rev.* 40(3):154–174.

Oestreicher-Singer G, Sundararajan A (2012) Recommendation networks and the long tail of electronic commerce. *MIS Quart.* 36(1):65–83.

Okoli C, Mehdi M, Mesgari M, Nielsen F, Lanamki A (2012) The people's encyclopedia under the gaze of the sages: A systematic review of scholarly research on Wikipedia. SSRN Working paper 2021326, Technical University of Denmark, Aarhus, Denmark.

Paolacci G, Chandler J, Ipeirotis PG (2010) Running experiments on Amazon Mechanical Turk. *Judgment Decision Making* 5(5): 411–419.

Podolny JM, Baron JN (1997) Resources and relationships: Social networks and mobility in the workplace. *Amer. Sociol. Rev.* 62(5):673–693.

Preece J, Schneiderman B (2009) The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Trans. Human-Comput. Interaction* 1(1):13–32.

Rad HS, Barbosa D (2012) Identifying controversial articles in Wikipedia: A comparative study. *Proc. Eighth Annual Internat. Sympos. Wikis Open Collaboration* (ACM, New York), 13–22.

Ransbotham S, Kane GC (2011) Membership turnover and collaborative success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quart.* 35(3):613–627.

Ransbotham S, Kane GC, Lurie N (2012) Network characteristics and the value of collaborative user-generated content. *Marketing Sci.* 31(3):387–405.

Raymond ES (1999) *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (O'Reilly, Cambridge, MA).

Scott J (2012) *Social Network Analysis*, 3rd ed. (Sage Publications, Thousand Oaks, CA).

Singh PV, Tan Y (2011) Developer heterogeneity and formation of communication networks in open source software projects. *J. Management Inform. Systems* 27(3):179–210.

Singh PV, Tan Y, Mookerjee V (2011a) Network effects: The influence of structural capital on open source project success. *MIS Quart.* 35(4):813–829.

Singh PV, Tan Y, Youn N (2011b) A hidden Markov model of developer learning dynamics in open source software projects. *Inform. Systems Res.* 22(4):790–807.

Stvilia B, Twidale MB, Smith LC, Gasser L (2008) Information quality work organization in Wikipedia. *J. Amer. Society Inform. Sci. Tech.* 59(6):983–1001.

Surowiecki J (2004) *The Wisdom of Crowds* (Doubleday, New York).

Trier M (2008) Towards dynamic visualization for understanding evolution of digital communication networks. *Inform. Systems Res.* 19(3):335–350.

Visser I, Speekenbrink M (2010) Depmixs4: An *R*-package for hidden Markov models. *J. Statistical Software* 36(7):1–21.

Wang XQ, Butler BS, Ren YQ (2013) The impact of membership overlap on growth: An ecological competition view of online groups. *Organ. Sci.* 24(2):414–431.

Warncke-Wang M, Cosley D, Riedl J (2013) Tell me more: An actionable quality model for Wikipedia. *Proc. 9th Internat. Sympos. Open Collaboration* (ACM, New York), Article 8.

Wilkinson DM, Huberman BA (2007) Cooperation and quality in Wikipedia. *Proc.* 2007 *Internat. Sympos. Wikis* (ACM, New York), 157–164.

Xu Y, Luo T (2011) Measuring article quality in Wikipedia: Lexical clue model. 2011 *3rd Sympos. Web Soc. (SWS)* (IEEE, Piscataway, NJ), 141–146.

Zhang X, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.

Zhu H, Kraut R, Kittur A (2012) Effectiveness of shared leadership in online communities. *Proc. ACM* 2012 *Conf. Comput. Supported Cooperative Work* (ACM, New York), 407–416.